DOCUMENT RESUME

ED 287 903 TM 870 692

AUTHOR Halpin, Gerald; And Others
TITLE Reliability of Ratings of Research Proposals Submitted for Funding.
PUB DATE Nov 86
NOTE 11p.; Papers presented at the Annual Meeting of the Mid-South Educational Research Association (Memphis, TN, November 19-21, 1986). For a related document, see TM 870 693.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; Evaluation Criteria; Evaluators; *Grants; Higher Education; *Interrater Reliability; *Peer Evaluation; *Reliability; *Research Proposals
IDENTIFIERS *Review Panels

ABSTRACT

Based upon the assumption that the process of peer review of publications and research is flawed, interrater reliability of reviews of 188 research proposals submitted for funding at a major university was studied. The eight dimensions rated were: (1) significance of the research; (2) clarity and reasonableness of the objectives; (3) appropriateness of the methodology; (4) adequacy and clarity of the budget; (5) potential for future extramural support; (6) applicant's experience; (7) review of the related literature; and (8) consistency of proposed research with the applicant's educational background and experience. Reliability coefficients across application periods, raters, and dimensions were obtained. The results of this study showed that evaluators of proposals submitted for funding within a university were not likely to agree on their ratings within and across application periods. Brief suggestions regarding the future training of evaluators were made. (MDE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED287903

Reliability of Ratings of Research Proposals Submitted for Funding

Gerald Halpin, Glennelle Halpin, Rhonda Blood

Auburn University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Glennelle Halpin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the meeting of the Mid-South Educational Research Association, Memphis, Tennessee, November 1986.

TM 870 692

BEST COPY AVAILABLE

Reliability of Ratings of Research Proposals Submitted for Funding

The process of peer review controls access to publication and money in modern academia (Horrobin, 1982). That the process is flawed has been suggested over the years.

In one of the earlier controlled experimental studies on interrater reliability and reviewer bias, Mahoney (1977) found that reviewers showed little interrater agreement on specifically scored components of the research article. Peters and Ceci (1982) found that reviewer bias was more significant than objective ratings of research quality in the professional journal peer review process. Other writers support the thesis that review for publication is a "noisy" process, often full of emotional responses and unsubstantiated judgments (Spencer, Hartnett, & Majoney, 1985). The prevailing impression is that the probability of a manuscript's publication depends more on luck and editorial or reviewer bias than on quality (Whitehurst, 1984).

Further support of the inadequacy of review procedures was provided by Cole, Cole, and Simon (1981). As part of their extensive analysis of the review system of the National Science Foundation for awarding research grants, they found that the reviewers themselves contributed substantially more variance than did the research proposals. They concluded that the fate of a proposal depended more heavily upon the particular reviewers who

happened to be selected than upon the merits of the proposal itself.

Others have carried out studies and reported similar results (Bowen, Perloff, & Jacoby, 1972; Gottfredson, 1978; Scott, 1974; Ward, Hall, & Schramm, 1975). Their findings have been fairly consistent:

1. Reviewers of research articles and funding proposals account for more variability than do any other objectively measurable factors.

2. A significant proportion of the variability (50%-70% or more) is due to chance.

Would the findings be different in an internal evaluation with peer evaluators rating peer proposals? This study was designed to answer that question.

Specifically, the primary objective of this study was to determine if established researchers at a major state university could reliably evaluate on eight dimensions of research proposals submitted by their peers for funding. A further objective was to determine if the reliability of these judgments was consistent from application period to application period with proposals and evaluators changing.

Method

Faculty members submitted research proposals to the university grant-in-aid committee for evaluative consideration for

possible funding. The number submitting proposals was 37, 35, 39, 39, and 38, respectively, over five application periods. Serving on the grant-in-aid committee for each respective application period were 7, 8, 7, 9, and 10 other faculty members who, themselves, were recognized researchers at the university.

Working independently, each committee member rated each proposal on eight dimensions: significance of the research, clarity and reasonableness of the objectives, appropriateness of the methodology, adequacy and clarity of the budget, potential for future extramural support, applicant's experience, review of the related literature, and consistency of proposed research with the applicant's educational background and experience. The maximum weights possible were 25, 10, 10, 10, 10, 15, 10, and 10 for each respective dimension. Ratings on each of these dimensions were summed to yield a total for each proposal.

For each of the five application periods, the ratings on each of the evaluated dimensions were subjected to a two-factor (proposal and rater) repeated measures analysis of variance without replications. Each respective between proposals mean square and residual mean square were used in the computation of average interrater reliability coefficients interpreted as a reliability estimate for an individual rater. Then the Spearman-Brown prophesy formula was applied to each of the 45 individual coefficients to obtain an estimate of the average interrater

reliability for four raters (typically the number of doctoral committee members) and for eight raters (the median number of raters across the five application periods in this study).

Results

Reliability coefficients across application periods, raters, and dimensions are given in Table 1.

Insert Table 1 about here

The range of reliability coefficients across applications for one rater, four raters, and eight raters respectively was .09/.25, .28/.57, .43/.72 for significance; .05/.11, .18/.32, .31/.49 for objectives; .07/.15, .25/.41, .40/.58 for methodology; .12/.33, .36/.66, .53/.79 for budget; .15/.31, .42/.64, .59/.78 for extramural support; .28/.69, .61/.81, .77/.86 for experience; .01/.18, .05/.47, .09/.64 for review of literature; .03/.20, .12/.51, .22/.67 for research background, and .20/.32, .49/.65, and .66/.79 for total.

Discussion

Results of this study show that evaluators of proposals submitted for funding within a university are not likely to agree on their ratings within and across application periods. More positive findings of importance to evaluators in all contexts emerged. First, as the number of raters increased, the

reliability of the ratings tended to increase. Second, the reliability of the total tended to be higher than the reliability of ratings for each separate dimension. Third, the reliability of the ratings of some of the dimensions was higher than the reliability of the ratings of other dimensions.

Future research wherein evaluators are trained to a criterion level of interrater agreement before evaluating research proposals might yield more positive findings. Lindsey (1976) targeted the review process as the culprit. He asserted that reviewers are unaware of the extent to which personal bias influences cognitive processes and that the integrity of reviewers is not in question. Perhaps more sensitivity to the lack of interrater agreement and improvements in the review process could reverse the prevailing negative evaluation of evaluation.

References

Bowen, D. D., Perloff, R., & Jacoby, J. (1972). Improving manuscript evaluation procedures. American Psychologist, 27, 221-225.

Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. Science, 214, 881-886.

Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. American Psychologist, 33, 920-934.

Horrobin, D. F. (1982). Peer review: A philosophically faulty concept which is proving disastrous for science. In S. Harnad (Ed.), Peer commentary on peer review (pp. 33-34). Cambridge, MA: Cambridge University Press.

Lindsey, D. (1978). The scientific publication system in social science. San Francisco: Jossey-Bass.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. Cognitive Therapy and Research, 1, 161-175.

Peters, D. P., & Ceci, S. J. (1982). Peer review practices of psychological journals: The fate of published articles, submitted again. In S. Harnad (Ed.), Peer commentary on peer review (pp. 3-11). Cambridge, MA: Cambridge University Press.

Scott, W. A. (1974). Inter-referee agreement on some characteristics of manuscripts submitted to the Journal of Personality and Social Psychology. American Psychologist, 29, 698-702.

Spencer, N. J., Hartnett, J., & Mahoney, J. (1985). Problems with reviews in the standard editorial practice. Journal of Social Behavior and Personality, 1, 21-36.

Ward, A. W., Hall, B. W., & Schramm, C. F. (1975). Evaluation of published educational research: A national survey. American Educational Research Journal, 12, 109-128.

Whitehurst, G. J. (1984). Interrater agreement for journal manuscript reviews. American Psychologist, 39, 22-28.

Table 1

Reliability Coefficients Across Application Periods, Raters, and Dimensions

Number of raters	Dimensions								
	Significance	Objectives	Methodology	Budget	Support	Experience	Literature	Background	Total
				Period 1					
1	.19	.11	.13	.33	.27	.35	.01	.16	.24
4	.48	.32	.37	.66	.60	.69	.05	.43	.55
8	.65	.49	.54	.79	.75	.81	.09	.60	.71
				Period 2					
1	.25	.07	.15	.16	.31	.28	.11	.03	.32
4	.57	.24	.41	.43	.64	.61	.34	.12	.65
8	.72	.38	.58	.60	.78	.77	.50	.22	.79
				Period 3					
1	.12	.10	.07	.16	.15	.69	.06	.06	.20
4	.35	.30	.25	.43	.42	.81	.21	.19	.49
8	.52	.47	.40	.61	.59	.85	.34	.32	.66

(table continues)

Table 1 (Continued)

Number of raters	Dimensions								
	Significance	Objectives	Methodology	Budget	Support	Experience	Literature	Background	Total
				Period 4					
1	.09	.05	.08	.12	.19	.43	.18	.19	.26
4	.28	.18	.26	.36	.48	.75	.47	.49	.58
8	.43	.31	.41	.53	.65	.86	.64	.66	.74
				Period 5					
1	.18	.06	.08	.20	.31	.40	.17	.20	.25
4	.47	.21	.25	.50	.64	.73	.44	.51	.57
8	.64	.34	.40	.66	.78	.84	.61	.67	.72

Note. Period 1: Number of applicants = 37; number of raters = 7.

Period 2: Number of applicants = 35; number of raters = 8.

Period 3: Number of applicants = 39; number of raters = 7.

Period 4: Number of applicants = 39; number of raters = 9.

Period 5: Number of applicants = 38; number of raters = 10.